

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
Please do not report the images to the
Image Problem Mailbox.

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-015854

(43)Date of publication of application : 22.01.1999

(51)Int.Cl.

G06F 17/30

(21)Application number : 09-181818

(71)Applicant : OMRON CORP

(22)Date of filing : 24.06.1997

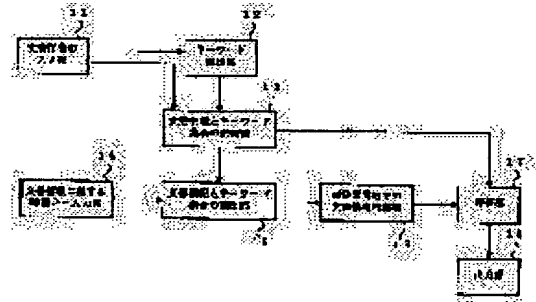
(72)Inventor : SENOO YOSHINOBU
KITAMURA MASAZUMI
FUJII FUJIKI
GO ATOU

(54) DOCUMENT PROCESSING METHOD, ITS DEVICE AND RECORDING MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a document processor capable of changing knowledge data following a change in user's interesting object and executing precise judgement.

SOLUTION: A keyword extraction part 12 extracts a keyword and its importance from an applied document and stores the extracted contents in a storage part 13 together with the preparation time of the document. A document information/keyword set extraction part 15 extracts a keyword and its importance in a document existing in an object section specified by optional time on a time base applied by a time key input means 14 front the storage part 13, generates knowledge data in the object section based on the extracted keyword and importance and stores the generated data in a document information storing part 16 in the object section. When an unknown document is applied, an evaluation part 17 compares a keyword and importance extracted from the unknown document with knowledge data stored in the storage part 16 to evaluate the unknown document.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision]

of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-15854

(43) 公開日 平成11年(1999) 1月22日

(51) Int.Cl.⁶
G 0 6 F 17/30

識別記号

F I
G 0 6 F 15/401 3 2 0 Z
15/40 3 7 0 A
15/401 3 1 0 A
15/403 3 4 0 A

審査請求 未請求 請求項の数21 F D (全 14 頁)

(21) 出願番号 特願平9-181818

(22) 出願日 平成9年(1997) 6月24日

(71) 出願人 000002945

オムロン株式会社

京都府京都市右京区花園土堂町10番地

(72) 発明者 妹尾 吉紳

京都府京都市右京区花園土堂町10番地 オムロン株式会社内

(72) 発明者 北村 正純

京都府京都市右京区花園土堂町10番地 オムロン株式会社内

(72) 発明者 藤居 藤樹

京都府京都市右京区花園土堂町10番地 オムロン株式会社内

(74) 代理人 弁理士 松井 伸一

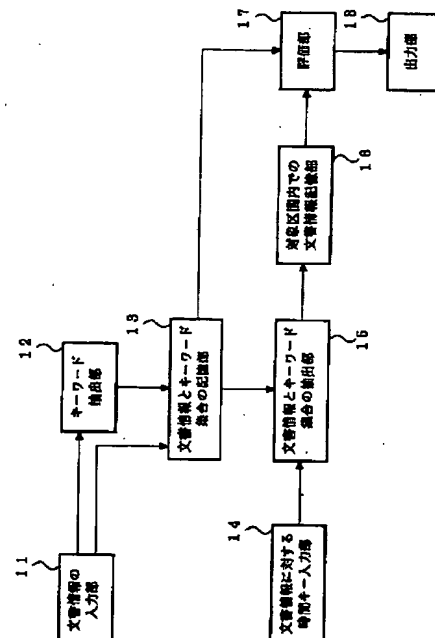
最終頁に続く

(54) 【発明の名称】 文書処理方法及び装置並びに記録媒体

(57) 【要約】

【課題】 ユーザーの興味対象等が変化してもそれに追従して知識データを変更し、精度のよい判断を行うことができる文書処理装置を提供すること

【解決手段】 与えられた文書中からキーワードとその重要度をキーワード抽出部12で抽出し、その文書の作成時間とともに記憶部13に格納する。文書情報とキーワード集合の抽出部15では、時間キー入力手段14により与えられた時間軸上の任意の時刻により特定される対象区間内に存在する文書についてのキーワードと重要度を、記憶部13から抽出するとともに、その抽出したキーワードと重要度に基づいてその対象区間における知識データを生成し、対象区間内の文書情報記憶部16に格納する。そして、評価部17では、未知の文書が与えられた場合に、その未知の文書から抽出されたキーワード及び重要度と、記憶部16に格納された知識データとを比較し、未知の文書に対する評価を行う。



【特許請求の範囲】

【請求項 1】 与えられた文書中からキーワードを抽出し、

前記文書内での前記キーワードの重要度を算出し、
前記抽出されたキーワード及び重要度と、前記与えられた文書の作成時間情報を含むデータを第 1 記憶手段に格納し、

時間軸上の任意の時刻を入力することにより設定される対象区間に基いて前記第 1 記憶手段にアクセスし、前記第 1 記憶手段に格納されたデータのうち、前記文書の作成時間が前記対象区間内に存在する文書についてのキーワードと重要度を抽出するとともに、その抽出したキーワードと重要度に基づいてその対象区間における知識データを生成し、

その生成された前記対象区間内での知識データを第 2 記憶手段に格納することを特徴とする文書処理方法。

【請求項 2】 未知の文書が与えられた場合に、その未知の文書からキーワード及び重要度を抽出し、前記抽出した未知の文書のキーワードと重要度を前記第 2 記憶手段に格納された知識データと比較し、前記文書に対する評価を行うようにしたことを特徴とする請求項 1 に記載の文書処理方法。

【請求項 3】 前記文書を与えるに際し、カテゴリを関連付けて入力し、

前記第 1 記憶手段に格納するデータを前記カテゴリごとに分けて格納し、

かつ、前記カテゴリごとに前記知識データを生成するとともに、前記第 2 記憶手段にカテゴリごとに格納することを特徴とする請求項 1 に記載の文書処理方法。

【請求項 4】 未知の文書が与えられ場合に、その未知の文書からキーワード及び重要度を抽出し、

その抽出したキーワード及び重要度と、前記第 2 記憶手段に格納されたカテゴリごとの知識データとを比較して各カテゴリに対する一致度を求め、

かつ、前記求めた各カテゴリに対する一致度に基づいて未知の文書を分類するようにしたことを特徴とする請求項 3 に記載の文書処理方法。

【請求項 5】 異なる対象区間を複数指定し、

指定された複数の対象区間ごとの知識データを生成し、各知識データの特徴を比較することでユーザーの興味の推移を求めるようにしたことを特徴とする請求項 1～4 のいずれか 1 項に記載の文書処理方法。

【請求項 6】 与えられた文書中からキーワードを抽出し、

前記文書内での構造レベルやキーワードの頻度に応じて前記キーワードの重要度を算出し、

前記抽出したキーワード及び重要度を累積演算によって過去に与えられた文書に基づいて生成されたキーワード及び重要度に足し合わせることににより知識データを生成し、

その生成した知識データを記憶手段に記憶させるようにし、

かつ前記累積演算は、与えられた文書中から抽出されたキーワード及び重要度の影響が大きくなるようにしたことを特徴とする文書処理方法。

【請求項 7】 与えられた未知の文書からキーワード及び重要度を抽出し、

その抽出した前記キーワード及び重要度と前記記憶手段に格納された累積された知識データとを比較し、前記文書に対する評価を行うようにしたことを特徴とする請求項 6 に記載の文書処理方法。

【請求項 8】 与えられた文書中からキーワードを抽出するとともに、前記文書内での前記キーワードの重要度を算出するキーワード抽出手段と、

前記キーワード抽出手段で抽出されたキーワードと、前記与えられた文書の作成時間情報を含むデータを記憶する第 1 記憶手段と、

時間軸上の任意の時刻を入力する時間キー入力手段と、前記第 1 記憶手段に格納されたデータのうち、前記文書の作成時間が前記時間キー入力手段から与えられる時刻により特定される前記対象区間内に存在する文書についてのキーワードと重要度を抽出するとともに、その抽出したキーワードと重要度に基づいてその対象区間における知識データを生成する生成手段と、

その生成手段により生成された前記対象区間内での知識データを記憶する第 2 記憶手段とを備えたことを特徴とする文書処理装置。

【請求項 9】 未知の文書が与えられた場合に、その未知の文書から抽出されたキーワード及び重要度と、前記第 2 記憶手段に格納された知識データとを比較し、前記文書に対する評価を行う評価手段をさらに備えたことを特徴とする請求項 8 に記載の文書処理装置。

【請求項 10】 前記文書を与えるに際し、カテゴリを関連付けて入力する機能を有し、

前記第 1 記憶手段に格納するデータは、前記カテゴリごとに分けて格納するようにし、

前記生成手段では、前記カテゴリごとに前記知識データを生成し、前記第 2 記憶手段にカテゴリごとに格納するようにしたことを特徴とする請求項 8 に記載の文書処理装置。

【請求項 11】 未知の文書が与えられた場合に、その未知の文書から抽出されたキーワード及び重要度と、前記第 2 記憶手段に格納されたカテゴリごとの知識データとを比較し、各カテゴリに対する一致度に基づいて評価した文書を分類するようにしたことを特徴とする請求項 10 に記載の文書処理装置。

【請求項 12】 対象区間を複数指定することにより、知識データを複数個生成し、

各知識データの特徴を比較することでユーザーの興味の推移を求める機能を備えたことを特徴とする請求項 8～

1 1 のいずれか 1 項に記載の文書処理装置。

【請求項 1 3】 与えられた文書中からキーワードを抽出するとともに、前記文書内での前記キーワードの重要度を算出するキーワード抽出手段と、
前記キーワード抽出手段で抽出されたキーワード及び重要度を累積演算によって過去に与えられた文書に基づいて生成されたキーワード及び重要度に足し合わせることに
より知識データを生成する生成手段と、
その生成手段により生成された知識データを記憶する記憶手段とを備え、

かつ前記生成手段における累積演算は、キーワード抽出手段で抽出されたキーワード及び重要度の影響が大きくなるようにしたものであることを特徴とする文書処理装置。

【請求項 1 4】 未知の文書が与えられた場合に、その未知の文書から抽出されたキーワード及び重要度と、前記記憶手段に格納された累積された知識データとを比較し、前記文書に対する評価を行う評価手段をさらに備えたことを特徴とする請求項 1 3 に記載の文書処理装置。

【請求項 1 5】 与えられた文書中からキーワードを抽出する処理、

前記文書内での前記キーワードの重要度を算出する処理、

前記抽出されたキーワード及び重要度と、前記与えられた文書の作成時間情報を含むデータをコンピュータ内の第 1 記憶手段に格納する処理、

時間軸上の任意の時刻を入力することにより設定される対象区間に基づいて前記第 1 記憶手段をアクセスし、前記第 1 記憶手段に格納されたデータのうち、前記対象区間内に存在する文書についてのキーワードと重要度を抽出するとともに、その抽出したキーワードと重要度に基づいてその対象区間における知識データを生成する処理、

その生成された前記対象区間内での知識データを前記コンピュータ内の第 2 記憶手段に格納する処理、

をコンピュータに実行させる命令を含む文書処理プログラムを格納したコンピュータが読取り可能な記録媒体。

【請求項 1 6】 与えられた未知の文書からキーワード及び重要度を抽出する処理、

前記抽出した未知の文書のキーワードと重要度を前記第 2 記憶手段に格納された知識データとを比較し、前記文書に対する評価を行う処理をさらに含むプログラムを格納した請求項 1 5 に記載の記録媒体。

【請求項 1 7】 前記第 1 記憶手段に格納するデータを前記カテゴリごとに分けて格納する処理、
前記カテゴリごとに前記知識データを生成するとともに、前記第 2 記憶手段にカテゴリごとに格納する処理をさらに含むプログラムを格納した請求項 1 5 に記載の記録媒体。

【請求項 1 8】 未知の文書が与えられた場合に、その

未知の文書からキーワード及び重要度を抽出する処理、その抽出したキーワード及び重要度と、前記第 2 記憶手段に格納されたカテゴリごとの知識データとを比較し、各カテゴリに対する一致度を求める処理、前記求めた各カテゴリに対する一致度に基づいて未知の文書を分類する処理をさらに含むプログラムを格納した請求項 1 7 に記載の記録媒体。

【請求項 1 9】 指定された複数の対象区間ごとの知識データを生成する処理、

10 各知識データの特徴を比較することでユーザーの興味の推移を求める処理をさらに含むプログラムを格納した請求項 1 5 に記載の記録媒体。

【請求項 2 0】 与えられた文書中からキーワードを抽出する処理、

前記文書内での構造レベルやキーワードの頻度に応じて前記キーワードの重要度を算出する処理、

前記抽出したキーワード及び重要度の影響が大きくなるように、累積演算によって過去に与えられた文書に基づいて生成されたキーワード及び重要度に足し合わせることに
より知識データを生成する処理、

20 その生成した知識データを記憶手段に記憶させる処理、
をコンピュータに実行させる命令を含む文書処理プログラムを格納したコンピュータが読取り可能な記録媒体。

【請求項 2 1】 与えられた未知の文書からキーワード及び重要度を抽出する処理、

その抽出した前記キーワード及び重要度と前記記憶手段に格納された累積された知識データとを比較し、前記文書に対する評価を行う処理をさらに含むプログラムを格納した請求項 2 0 に記載の記録媒体。

30 【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】本発明は、文書処理方法及び装置並びに記録媒体に関するもので、より具体的には、与えられた文書からキーワードを抽出し、その文書の特徴を求めて評価基準となる知識データを生成したり、その生成した知識データに基づいて未知の文書を評価し、知識データに対する一致度などを求めることにより、ユーザーの興味のあるものか否かや内容の推定を行うことのできる文書処理を行うものに関する。

40 【0 0 0 2】

【従来の技術】従来、文書特徴量を用いた文書処理は一般化しており、例えば、文書検索の分野においては、問合わせ式と文書との一致度を計算する尺度として、問合わせキーワードの文書中での出現頻度（キーワードの重要度を表わす）やキーワードの出現する文書数（キーワード弁別能力を表わす）を用いている。これにより、多数の文書から目的とする内容が記載された文書を抽出したり、新規の未知文書があるカテゴリに属する文書であるか否かのおおよその判断ができるようになってい

50 る。

【0003】

【発明が解決しようとする課題】ところで、新規文書の内容が、ユーザーにとって興味のある文書であるか否かなどを判断することは、例えば電子メール文書等において多数のメールが届くような場合にその電子メールの内容を予めコンピュータ側で解析しユーザーにとって興味のある（見る）必要のある文書であるか否かを判断することができると、メールを見るか否かの判断の指標となり、効率よくメールに対する処理ができて好ましい。

【0004】そこで、精度よく判断する必要があるが、上記した判断の精度は、比較基準となる知識データの良否が大きな要因となる。そして、ユーザーにとって見る必要がある／興味の有する文書は、時間の経過とともに変化するのが通常である。したがって、ある時期で有効に判断できた知識も、時間経過とともに有効に判断できなくなるおそれがある。したがって、知識データをメンテナンスして常にユーザーにとって適した状態にすることが必要であるが、個々のユーザーで興味は異なるので、実際に個々のユーザーに合わせてメンテナンスをすることは煩雑で、実用に供し得ない。

【0005】本発明は、上記した背景に鑑みてなされたもので、その目的とするところは、上記した問題を解決し、従来の文書特徴量に加え、時間的要素を加味することにより、ユーザーの興味対象等が変化してもそれに追従して知識データを変更し、精度のよい判断を行うことができ、また、ユーザーの興味対象やその逆が時間とともにどのように推移していったかを検出することのできる文書処理方法及び装置並びに記録媒体を提供することにある。

【0006】

【課題を解決するための手段】上記した目的を達成するため、本発明に係る文書処理方法では、まず第1の発明（第1の実施の形態で実現されている）として、与えられた文書中からキーワードを抽出し、前記文書内での前記キーワードの重要度を算出し、前記抽出されたキーワード及び重要度と、前記与えられた文書の作成時間情報を含むデータを第1記憶手段に格納する。ここまでの工程は、実施の形態では図3に示すフローにより実現される。次いで、時間軸上の任意の時刻を入力することにより設定される対象区間に基づいて前記第1記憶手段をアクセスし、前記第1記憶手段に格納されたデータのうち、前記文書の作成時間が前記対象区間内に存在する文書についてのキーワードと重要度を抽出するとともに、その抽出したキーワードと重要度に基づいてその対象区間における知識データを生成し、その生成された前記対象区間内の知識データを第2記憶手段に格納することができる（請求項1）。ここまでの工程は実施の形態では図4に示すフローにより実現される。

【0007】上記の方法を前提とし、未知の文書が与えられた場合に、その未知の文書からキーワード及び重要

度を抽出し、前記抽出した未知の文書のキーワードと重要度を前記第2記憶手段に格納された知識データと比較し、前記文書に対する評価を行うようにしてもよい（請求項2）。この工程は、実施の形態では、図7により実現されている。

【0008】前記文書を与えるに際し、カテゴリーを関連付けて入力し、前記第1記憶手段に格納するデータを前記カテゴリーごとに分けて格納し、かつ、前記カテゴリーごとに前記知識データを生成するとともに、前記第2記憶手段にカテゴリーごとに格納するようにしてもよい（請求項3）。

【0009】未知の文書が与えられた場合に、その未知の文書からキーワード及び重要度を抽出し、その抽出したキーワード及び重要度と、前記第2記憶手段に格納されたカテゴリーごとの知識データとを比較し、各カテゴリーに対する一致度を求め、かつ、前記求めた各カテゴリーに対する一致度に基づいて未知の文書を分類するようにしてもよい（請求項4）。

【0010】そして、上記した各方法を実施するために適した第1の発明としては、例えば与えられた文書中からキーワードを抽出するとともに、前記文書内での前記キーワードの重要度を算出するキーワード抽出手段と、前記キーワード抽出手段で抽出されたキーワードと、前記与えられた文書の作成時間情報を含むデータを記憶する第1記憶手段（実施の形態では、「文書情報とキーワード集合の記憶部13」に対応）と、時間軸上の任意の時刻を入力する時間キー入力手段と、前記第1記憶手段に格納されたデータのうち、前記文書の作成時間が前記時間キー入力手段から与えられる時刻により特定される前記対象区間内に存在する文書についてのキーワードと重要度を抽出するとともに、その抽出したキーワードと重要度に基づいてその対象区間における知識データを生成する生成手段と、その生成手段（実施の形態では、「文書情報とキーワード集合の抽出部15」に対応）により生成された前記対象区間内の知識データを記憶する第2記憶手段（実施の形態では「対象区間内の文書情報記憶部16」に対応）とを備えて構成することができる（請求項8）。

【0011】また、未知の文書が与えられた場合に、その未知の文書から抽出されたキーワード及び重要度と、前記第2記憶手段に格納された知識データとを比較し、前記文書に対する評価を行う評価手段をさらに備えるように構成してもよい（請求項9）。

【0012】また、前記文書を与えるに際し、カテゴリーを関連付けて入力するようにし、前記第1記憶手段に格納するデータは、前記カテゴリーごとに分けて格納するようにし、前記生成手段では、前記カテゴリーごとに前記知識データを生成し、前記第2記憶手段にカテゴリーごとに格納するように構成してもよい（請求項10）。

【0013】さらにまた、未知の文書が与えられた場合に、その未知の文書から抽出されたキーワード及び重要度と、前記第2記憶手段に格納されたカテゴリーごとの知識データとを比較し、各カテゴリーに対する一致度に基づいて評価した文書を分類する（実施の形態では、「評価部17」により実現される）ように構成してもよい（請求項11）。

【0014】上記した目的を達成するための第2の発明（第3の実施の形態として実現されている）としては、上記した請求項1～4を前提とし、異なる対象区間を複数指定し、指定された複数の対象区間ごとの知識データを生成し、各知識データの特徴を比較することでユーザーの興味の推移を求めるようにすることである（請求項5）。

【0015】そして、係る方法を実施するための装置としては、請求項8～11に記載の装置を前提とし、対象区間を複数指定することにより、知識データを複数個生成し、各知識データの特徴を比較することでユーザーの興味の推移を求める機能（実施の形態では、「ユーザーの興味推移算出部19」に対応）をさらに備えることである（請求項12）。

【0016】上記した目的を達成するための第3の発明（第2の実施の形態として実現されている）として、与えられた文書中からキーワードを抽出し、前記文書内での構造レベルやキーワードの頻度に応じて前記キーワードの重要度を算出し、前記抽出したキーワード及び重要度を累積演算によって過去に与えられた文書に基づいて生成されたキーワード及び重要度に足し合わせることで知識データを生成し、その生成した知識データを記憶手段に記憶させるようにし、かつ前記累積演算は、与えられた文書中から抽出されたキーワード及び重要度の影響が大きくなるようにすることである（請求項6）。

【0017】上記方法を前提とし、与えられた未知の文書からキーワード及び重要度を抽出し、その抽出した前記キーワード及び重要度と前記記憶手段に格納された累積された知識データとを比較し、前記文書に対する評価を行うようにしてもよい（請求項7）。

【0018】そして、係る方法を実施するための装置としては、与えられた文書中からキーワードを抽出するとともに、前記文書内での前記キーワードの重要度を算出するキーワード抽出手段と、前記キーワード抽出手段で抽出されたキーワード及び重要度を累積演算によって過去に与えられた文書に基づいて生成されたキーワード及び重要度に足し合わせることで知識データを生成する生成手段と、その生成手段により生成された知識データを記憶する記憶手段とを備え、かつ前記生成手段における累積演算は、キーワード抽出手段で抽出されたキーワード及び重要度の影響が大きくなるように構成した（請求項13）。

【0019】そして、上記装置を前提とし、与えられた

未知の文書から抽出されたキーワード及び重要度と、前記記憶手段に格納された累積された知識データとを比較し、前記文書に対する評価を行う評価手段をさらに備えて構成するとよい（請求項14）。

【0020】さらに請求項15～21に規定するように、上記した各処理を実行するためのプログラムを格納した記録媒体としてもよい。

【0021】従来からある文書特徴量に加え、キーワードの持つ時間的、意味的な性質を考慮し、時系列的にキーワード集合をとらえることにより、ユーザーの興味対象あるいは、その逆となる文書、または、現在、話題の中心となっている文書あるいは、めずらしい話題の文書かを表わすことができる。

【0022】そして、請求項1, 3, 6, 8, 10, 13, 15, 17, 20の発明では、時間推移に基づくユーザーの興味の変化などに追従した文書の特徴量（知識データ）を生成できる。その結果、請求項2, 4, 7, 9, 11, 14, 16, 18, 21の発明では、上記した知識データに基づいて未知の文書に対する評価を精度よく行うことができ、ユーザーにとって必要な文書であるか否かの判断も容易に行うことができる。

【0023】そして、請求項6, 7, 13, 14, 20, 21のようにすると、知識データの記憶容量を削減できる。さらに、現在のキーワード集合と、過去のある時点でのキーワード集合との比較により、ユーザーの興味や話題の推移をも表現することができる。さらには、近い将来のユーザーの興味や話題の予測も可能となる（請求項5, 12, 19）。

【0024】*用語の定義

請求項1, 8等で規定される「対象区間」は、例えば2つの時間を指定することによりその間を特定するようにしてもよく、或いは、ある時間から最近まで（最初からある時間まで）等のように時間指定は1つでもよく、さらには、ある時間（時刻）というように「時」も含む概念である。

【0025】キーワード抽出手段は、キーワードを抽出する機能と重要度を求める機能をそれぞれ別々の処理部で実現するようにしてもよく、係る場合には、それら複数の処理部全体で本発明というキーワード抽出手段となる。

【0026】

【発明の実施の形態】まず、本発明の実施の形態の具体的な構成を説明する前に、概念的な実施の形態を説明する。一般にある文書がユーザーにとって興味対象となる文書であるか否かを自動的に判断する場合には、判断対象の文書中のキーワードを抽出し、そのキーワードを予め作成した辞書（キーワード）と比較して一致度を求め、その一致度の大小に基づいて判断することができる。

【0027】この時、まず文書中にあらわれるキーワー

ドを抽出する時において、時間軸をとり、その時点での時間と抽出されたキーワード集合を記憶する（図 1 参照）。そして、ある時点 A とある時点 B を指定し、その区間でのキーワードの重要度を累計することにより、ユーザーの興味対象となる文書、あるいは、その逆となる文書を表わすことができる。また、新規文書の読込み時に、過去のキーワード集合から、この文書が、ユーザーの興味対象となる文書あるいは、その逆である文書を表わすことができる。なお、以下に示す実施の形態では、ユーザーが受け取った文書を電子メール文書としているが、同様の処理が、その他の文書でも、可能である。

【0028】本形態では、受信した電子メール文書が、使用者にとって有意義な文書（興味対象となる文書）であるか否かを判断するための装置である。図 2 は本発明の好適な第 1 の実施の形態を示している。同図に示すように、まず入力側に文書情報の入力部 11 を設け、この文書情報の入力部 11 にて受信した電子メール文書情報を、次段のキーワード抽出部 12 及び文書情報とキーワード集合の記憶部 13 にそれぞれ与えるようになっている。また、ここで与える文書情報としては、電子メール文書自体のみでもよく、それに加えて文書情報の入力時間やその文書に対するユーザーからの情報（カテゴリ（興味の有無など））も併せて入力するようにしてもよい。なお、入力時間は、ユーザーがマニュアルで入力するようにしてもよいが、本発明では係る時間が重要な要素となるので、文書情報が入力された時に内蔵する時計をアクセスしてその時の時刻を取得し、それを関連づけて記憶部 13 に与えるようにしてもよい。また、入力時間は文書が電子メールとすると、その電子メールを受信した時とするのが最も好ましいが、システムの簡略化並びに通常の使用状況下では、電子メールの受信時刻と本装置への文書情報の入力時刻との差があまりないか、あったとしても一定の範囲内であるため、本装置への入力時刻を格納するようにしている。

【0029】キーワード抽出部 12 は、与えられた電子メール文書情報（電子メールを構成する文書）のキーワードを抽出するとともにその抽出したキーワードの重要度を求め、求めたキーワードと重要度を関連づけて文書情報とキーワード集合の記憶部 13 に送るようになっている。ここでキーワード抽出は、たとえば文書を分かち書きした後品詞解析を行って名詞のみを抽出し、それをキーワードとすることができる。また、重要度は、抽出したキーワード（名詞）の出現頻度等に基づいて求める（出現頻度が多いものほど重要度は高い）ことができる。なお、最終的に登録・保持するキーワードは、抽出されたもの総てでもよく、或いは求めた重要度が一定の値以上のものや上位 n 個のみを最終的なキーワードにするなど各種の方法が採れる。なお、係るキーワード抽出処理等は、文書をデータベース化する際に用いられる各

種の自動キーワード抽出法を適用できる。

【0030】文書情報とキーワード集合の記憶部 13 は、文書情報の入力時間を含めた文書情報およびキーワード抽出部 12 で抽出されたキーワード（重要度付）を関連づけて記憶する。なお、文書情報の入力部 11 により、ユーザーの興味やその逆を表わすカテゴリなどが指定されている場合には、記憶部 13 へはカテゴリごと分類して所定の情報を記憶する。

【0031】一方、入力手段として文書情報に対する時間キー入力部 14 を備えており、この時間キー入力部 14 は、新規文書に対して評価を行う際の知識データを生成する際の対象期間としての時間情報を入力するもので、その時間情報が次段の文書情報とキーワード集合の抽出部 15 に与えられる。この文書情報とキーワード集合の抽出部 15 は、与えられた時間情報に基づいて文書情報とキーワード集合の記憶部 13 にアクセスし、時間情報で設定される対象期間内に入力された文書に関する情報を抽出し、上記知識データを生成するようになっている。そして、生成した知識データ（キーワードの累計）を対象区間内の文書情報記憶部 16 に格納する。なお、この対象区間は、デフォルト値としては開始から現時点までの全区間としておく。

【0032】さらに本形態では、評価部 17 を有し、ここにおいて対象区間内の文書情報記憶部 16 に記憶されたキーワード集合と重要度の累計に基づき、新規に受信した電子メール文書の評価を行う。すなわち、この評価部 17 では、「評価対象の電子メール文書のキーワード（重要度付）」と、「文書情報記憶部 16 に記憶されたキーワード集合と重要度の累計」とを比較し、その一致度をみることによりユーザーにとって興味対象で見る必要が高いメールか否かを判断する。そこで、本形態では、評価対象の電子メール文書も文書情報の入力部 11 を介して与えられ、キーワード抽出部 12 でキーワードが抽出された後、文書情報とキーワード集合の記憶部 13 に記憶するようにしておく。そして、評価部 17 は、両記憶部 13、16 に格納された評価対象の電子メール文書のキーワード（重要度付）と、それまでに蓄積された文書データに基づいて生成された知識データ（キーワード集合と重要度の累計）を読み出し、比較処理して一致度を求める。

【0033】つまり、知識データが特にカテゴリ分類されることなく電子メール文書を与えたものに基づいて生成された場合、キーワード集合に属するキーワードと一致するものはユーザーにとって興味がある可能性が高い文書といえる（前提として、有益な興味のある文書をユーザーが登録する）。そして、キーワードの中でも重要度の累計が大きいものほど、興味を有する文書に存在する可能性の高いキーワードといえる。

【0034】従って、例えば評価対象の文書から抽出された重要度の高いキーワードが、知識データにおいても

キーワードとして抽出されしかも重要度の累計も大きい（一致度が高い）場合には、ユーザーが興味のある文書らしいと判断できる。逆に評価対象の文書から抽出された重要度の高いキーワードが、知識データになかったり仮にあったとしても重要度の累計が大きい場合には、ユーザーの興味のない文書らしいと判断できる（知識データにない方がより興味のないらしさが高くなる（一致度が低い））。さらにまた、仮に抽出されたキーワードと知識データに記憶されたキーワードが一致したとしても、例えば評価対象の文書において重要度の高いキーワードが知識データでは重要度の累計が小さく、評価対象の文書において重要度の低いキーワードが知識データでは重要度の累計が大きいというように、重要度の大小が逆になっているような文章も一致度は低くなる（なお、システム設計や実際のユーザーの使用などにあわせて、係の場合には一致度を高くするようにしてももちろんよい）。

【0035】そして、具体的な一致度の算出は、上記した事項に鑑み、キーワードと重要度（累計）を総合的に判断し、ファジィ推論その他の評価方法により求めてもよく、或いは単純に同一のキーワードの重要度（評価文書）と重要度の累計（過去の文書に基づく知識データ）とを乗算し、得られた値の総和を求め、その総和が大きいものほど一致度が高いとしたり、上記乗算する際に、片方にしかキーワードがない場合には、その値を0にしたり或いは減算対象にしたりするなど各種の方法が採れる。そして、本発明における評価部の処理としては、これらの演算処理に限られないのはもちろんである。

【0036】そして、その評価部17で得られた評価結果と電子メール文書およびその文書情報を、出力部18を介して出力するようになっている。この出力部18としてはCRTなどのディスプレイ（表示装置）であったり、プリンタなどの印刷装置であったりする。

【0037】次に上記構成の装置の使用法・作用を説明しつつ、各部の詳細な説明をする。

【0038】*文書情報収集工程

未知の文書进行评估するために必要な知識を生成する前準備として、受信した電子メール文書についての文書情報の収集を行う。具体的には、図3に示すように、まず文書情報の入力部11を用いて受信した文書メールの文書情報（文書+入力時間等）を入力する（ST1）。この入力処理は、通常の使用状況下では、電子メールを受信する都度、或いは電子メールをダウンロードしておき、一定の期間ごとに一括して入力するなど各種の方法が採れる。このようにすることにより、時間の経過とともに、情報が蓄積され、精度のよい評価・判定が行える。

【0039】次に、キーワード抽出部12にて与えられた文書情報のキーワードの出現頻度などにより、各キーワードごとの重要度を算出し、文書入力時点での時刻に対するキーワード集合を生成する。そして生成したキー

ワード集合と時刻を含む文書情報を記憶文書情報とキーワード集合の記憶部13に格納する（ST3）。これにより、情報収集工程が完了する。なお、ユーザーの興味あるいは、その逆を表わすカテゴリーを指定していれば、カテゴリーごとに文書情報記憶を生成する。そして、上記したステップ1～ステップ3は、新規文書を入力するごとに繰り返され、時系列的に文書情報とキーワード集合の記憶部13に記憶される。

【0040】*知識データ生成工程

この工程は、上記文書情報収集工程を経て記憶された文書情報から、未知の電子メールに対する評価を行う基準となる知識データを生成するものである。具体的には、図4に示すように、文書情報に対する時間キー入力部14を用いて時間軸上のある時点、あるいは、ある区間を指定する（ST5）。

【0041】その指定された区間をキーにして、文書情報とキーワード集合の記憶部13に記憶された情報をサーチし、該当する区間内に記憶された文書情報とキーワードを抽出する（ST6）。

【0042】次いで、抽出されたすべてのキーワードに付された重要度の累計を求め、その区間におけるキーワードとその求めた累計を関連づけて対象区間内の文書情報記憶部16に登録する（ST7）。この累計は、例えば同一のキーワードに付された重要度の総和を求めるとともに、その総和を抽出された文書の総数で除算するというように単純に平均値を求めてもよく、その他各種の方法を用いることができる。そして、係の累計の算出はすべてのキーワードについて行う。

【0043】この処理の一例を示すと、以下のようになる。まず、文書情報とキーワード集合の記憶部13の内部のデータ構造は、図5に示すように、文書を特定する文書名と、その電子メール文書を入力した日時と、キーワード（重要度付）を関連づけて格納しているものとする。なお、具体的な文書の内容もあわせて格納するようにしてもよいし、或いは他の記憶エリアに格納しておき、文書名などをキーにリンク付けをしてもよい。そして図では「A1, D1, D13」の3つの文書が示されているが、それ以外の文書も格納されている。

【0044】係の場合に、ステップ5で「1997/05/12 01:10～1997/05/30 00:00」までの区間を時間キーとして入力したとすると、図5中「時間」の欄をアクセスし、上記時間キーで特定される区間内に存在するものを抽出する。これにより、図6に示すように時間キーで設定された対象区間と、キーワード（重要度の累計）を関連づけて格納する。

【0045】*新規電子メール文書の評価工程

新規に電子メールを受信した場合に、図7に示すように、係の新規電子メール文書を文書情報の入力部14を用いて入力する（ST10）。次いでキーワード抽出部12が動作し、その文書からキーワードを抽出し、キー

ワード集合を生成する（ST11）。なお、ここまでの工程は、上記（1）におけるステップ1、2と同じである。

【0046】次に、評価部17が対象区間内での文書情報記憶部16にアクセスし、ステップ11で生成したキーワード集合（評価文書から抽出したキーワード集合（重要度付））をキーワード単位で比較し、その文書が、ユーザーの興味の対象となる文書、あるいは、その逆となる文書であるかの指標となる一致度を求める（ST12）。

【0047】そして、求めた一致度に基づき、その文書の分類を行う。すなわち、例えば評価対象の電子メール文書は、ユーザーにとって興味がわき必要のあるものか、逆に興味が無く見る必要のない文書である等を区分けする。そして、一致度とその分類をあわせて出力部8に出力する（ST13）。従って、ユーザーの興味対象あるいは、その逆となる文書であるかを内容を読む前に知ることができるので、メールを読むか否かの判断を簡単に行うことができる。

【0048】また、知識データを生成する際に文書とともにカテゴリーも入力していた場合には、知識データもカテゴリーごとに分離形成されるので、一致度の算出も各カテゴリーごとに行うことになる。そして、最も一致度の高いカテゴリーに属する可能性が高いと判断し、当該カテゴリーに属する文書と分類できる。

【0049】そのようにした場合には、例えばカテゴリーが仕事、趣味、プライベート等とすると、電子メールの内容のおおよその見当が付き、すぐに見た方がよいか後で見て大丈夫かの判断をしやすくなる。また、カテゴリー分類が、興味のある／ないとすると、どちらのカテゴリーの一致度が高いかにより評価対象の電子メール文書が、興味のあるものか否かを直接的に判断できる。しかも、どちらのカテゴリーに対する一致度も低い場合には、現在、話題の中心となっている文書あるいは、めずらしい話題の文書等である可能性が高いと判断することもできる。

【0050】また、適合するメールが多い場合、各属性ごとに出力表示することにより、分類して表示することができる。さらにまた、この実施の形態に検索機能を付加した場合、問い合わせキーワードへの重みづけにユーザーの現在の興味度などを加えることで、よりユーザー一人ひとりの時間的な興味の推移に応じた検索が可能となる。さらには、この実施の形態に要約機能を付加した場合についても、上記と同様によりユーザー一人ひとりの時間的な興味の推移に応じた要約が可能となる。

【0051】図8は、本発明の第2の実施の形態を示している。ユーザーが受け取った電子メール文書中のキーワード集合は、時間軸とともに変化する。そこで、上記した第1の実施の形態においては、文書ごとのキーワード集合すべてを記憶し、必要に応じて時間キーを入力し

てある時間区間におけるキーワード（重要度の累計付）を求め、知識を形成するようにしていた。そのため、開始からの経過時間が長くなると、記憶する文書数も多くなるため、文書情報とキーワード集合の記憶部13の記憶容量も大きなものが必要となる。

【0052】そこで本実施の形態においては、ユーザーが受け取った文書中のキーワード集合を随時累計していくことにより、単に、新規情報を追加することに比べ、記憶容量の削減を図ることができる。これらのキーワード集合において、頻繁に現れるキーワードについては、よく議論されているととらえることができる。また、時間とともに、しだいに現れなくなったキーワードについては、重要度を下げていくことより、忘却機能を与えることができる。

【0053】ここで具体的な構成の説明をする前に、概念（イメージ）を説明する。まず、新規電子メール文書が入力されるごとに、次の（1）と（2）の特徴をとらえ、その電子メールの文書が、どちらに属するかの評価を行い分類出力する。

- 20 （1）ユーザーに必要なあるいは、興味のある文書
- （2）不要あるいは、興味のない文書

また、随時、文書情報であるキーワード集合とその重要度を累計することにより、すべての文書ごとのキーワードを記憶する必要がなくなるので、記憶容量を小さくできるとともに、新規文書の特徴を次の文書に対する評価に反映することができる。

【0054】次に、具体的な構成について説明する。図8に示すように、新規の電子メール文書は文書情報の入力部21を介してキーワード抽出部22に与えられ、そこにおいてキーワードを抽出するとともに、出現頻度等を用いて各キーワードの重要度を求める。これら入力部21とキーワード抽出部22は、基本的に第1の実施の形態における入力部11とキーワード抽出部12と同様のものを用いることができる。

【0055】そして、キーワード抽出部22で抽出されたキーワード（重要度付き）は評価部23と知識更新部25に与えられる。この評価部23では文書情報記憶部24に格納された今までに与えられた文書情報から生成した知識データに基づいて評価対象の新規電子メール文書に対する評価（ユーザーの興味対象となるか否かなど）を行う。また知識更新部25では、次の評価に備えて新規文書について求めたキーワード集合と重要度および時刻を含む文書情報を、過去に記憶された文書情報に累計して新たな知識データを生成し、文書情報記憶部24を更新するようになっている。

【0056】さらに評価部23では、新規の文書について求めた評価結果、すなわち、ユーザーの興味の対象となる文書であるか否かを出力部25を用いて出力するようになっている。

- 50 【0057】次に、本実施の形態の要部となる知識更新

部26の機能について説明する。新規電子メール文書が
入力され、その文書中から抽出されたキーワードAの重
要度を下記の式に入力することで、文書情報記憶部に記*

文書情報記憶部に更新するキーワードAの重要度＝

$$\frac{\text{文書情報記憶部中の重要度} \times 0.9 + \text{新規電子メール文書中の重要度}}{1 + 0.9}$$

ここで、0.9とは、忘却係数を表わし、調整可能とする。

このことにより、新規電子メール文書で抽出されないキ
ーワードは、上記した式の右辺の分母の第2項が0とな
るので、文書情報記憶部中の重要度のほぼ半分が新規に
更新される際のそのキーワードの重要度となる。従っ
て、使用頻度が下がってきたあるいは、使われなくなっ
たキーワードに対する重要度は、小さくなっていく機
能、つまり、忘却機能を実現することが可能となる。ま
た、仮にキーワードとして抽出されていても、その抽出
された重要度の増減によって新規に更新される際のキ
ーワードの重要度も増減する。これにより、頻繁に使用さ
れるキーワードで、しかも最近受信した電子メール文書※20

{ペルー(0.4), 青木(0.7), 橋本(0.8), 帰国(0.4), ...}

すると、知識更新部25では、キーワード抽出部22か
ら出力される上記キーワード集合(重要度付)を受け取
るとともに、文書情報記憶部24にアクセスしてそこに
格納された現在の知識データを取得し、各キーワードに
対して、累計演算を施し、新たな重要度を求める。その
累計演算の一例を示すと、キーワード「ペルー」の場合
には、

【0061】

【数2】

$$\begin{aligned} \star \quad \text{「ペルー」の重要度} &= \frac{0.8 \times 0.9 + 0.4}{1 + 0.9} \\ &= 0.6 \end{aligned}$$

となる。

【0064】その他のキーワードも同様に計算すること
により、各キーワードについての重要度を求め、その求
めた新たなキーワード集合(重要度付)を文書情報記憶
部24に格納する。その結果、同図(B)に示すように
更新される。ここで、キーワード「人質」の重要度は下
がり、話題の中心となるキーワード「青木」の重要度が
上がる。なお、上記した説明では、評価部23と知識更
新部25を別構成にしたが、使用する情報(データ)は
同じであるので、1つの処理部で処理を行うように構成
してももちろんよい。

【0065】次に、上記した装置を用いた作用である本
発明に係る方法の実施の形態について説明する。前提と

*憶するキーワードAの重要度を求める。

【0058】

【数1】

10 ※中の重要度が増加する傾向にあるものは、最近興味が増
してきたものと判断でき、知識データとして格納される
キーワードの重要度も増すように作用する。

【0059】そして、具体例を示すと、まず、新規文書
入力直前の文書情報記憶部24に格納されたデータが、
図9(A)に示すようになっていくとする。次に、新規
文書が入力され、キーワード抽出部22により次のよう
なキーワード集合が生成されたとする。いずれも()の
中がその新規文書におけるキーワードの重要度である。

【0060】

【表1】

$$\begin{aligned} \star \quad \text{「ペルー」の重要度} &= \frac{0.8 \times 0.9 + 0.4}{1 + 0.9} \\ &= 0.6 \end{aligned}$$

となる。

30 【0062】同様に、新規文書でキーワードとして抽出
されなかった「フジモリ」の場合には、

【0063】

【数3】

$$\begin{aligned} \star \quad \text{「フジモリ」の重要度} &= \frac{0.6 \times 0.9 + 0}{1 + 0.9} \\ &= 0.3 \quad (\text{小数点以下2桁を四捨五入}) \end{aligned}$$

して、すでに過去に文書情報が入力されており、文書情
報記憶部24にはキーワード集合(重要度付)が登録さ
れているものとする。図11に示すフローチャートのよ
うに、まず、新規文書情報が入力される(ST21)。
これを受けてキーワード抽出部22では各キーワードに
対する重要度を出現頻度などにより算出し評価部23に
送る(ST22)。評価部23では、与えられた新規文
書についてのキーワード(重要度付)と過去に与えられ
た文書情報に基づいて生成された知識データを用いて、
キーワード単位で比較することにより、その新規文書が
ユーザーの興味の対象となる文書、あるいは、その逆と
なる文書であるかを評価する(ST23)。また、知識
更新部25にて、新規文書に基づくキーワード集合(重

要度付)を加味・累計して新たな知識データを生成し、文書情報の記憶部24に追加・更新する(ST24)。一方、上記の評価結果をもとに、ユーザーの興味ごとに、分類出力する(ST25)。

【0066】なお、上記した処理フローは、新規文書を入力するごとに随時繰り返され、文書情報の記憶部の知識データを更新する。

【0067】以上の構成により、随時、ユーザーの興味対象あるいは、その逆となる文書、または、現在、話題の中心となっている文書あるいは、めずらしい話題の文書かを表わすことができる。そして、文書情報記憶部には、新規情報を累計していくことにより、単に、新規情報を追加することに比べ、記憶容量が膨大になりすぎることを防ぐことができる。また、このことにより、文書情報記憶部へのアクセス処理速度の負荷を軽減すること*

(1) 先月の興味を表わすキーワード(重要度) :

旅行(0.8)、アメリカ(0.4)、ホテル(0.5)、スキー(0.9)

(2) 今月の興味を表わすキーワード(重要度) :

旅行(0.8)、日本(0.2)、ペンション(0.3)、スキー(0.9)

係る区間は、例えば文書情報に対する時間キー入力部14からユーザーが推移を見たいと思う2つの区間を指定することにより行える。また、例えば先月と先々月というように予め推移を見る区間を特定しておくことにより、自動的に入力するようにしてもよい。

【0071】推移算出部19では、比較する2つの区間のキーワード集合(重要度付)を取得し、文書検索技術分野において利用されるベクトル空間法によりそれぞれの区間についてのキーワードベクトルを求める。これにより、図12中(1)、(2)で示すようなグラフが求められるので、(1)のキーワードベクトルから(2)のキーワードベクトルにユーザーの興味が推移したことがわかる。そして、例えば係るグラフを出力部18に出力することができる。

【0072】また、上記のように2つの区間のみでなく、3つ以上の区間にてキーワード集合体を求めておき、各区間のキーワードベクトルを求めることにより、時間の経過によりどのようにユーザーの興味が推移していったかがより明確にわかり、さらには今後の推移の予測を行うことにより、例えば来月のキーワードベクトルの存在予測をし、対象区間内での文書情報記憶部16に予測したキーワード集合(重要度付)を格納し、評価部17における評価は、係る予測した知識データ(キーワード集合)に基づいて行うようにすることもできる。

【0073】以上、対象区間同士でのキーワード集合を比較することにより、ユーザーの興味あるいは、その逆の時間的な推移を抽出することができ、ユーザーの話題の時間的な推移を抽出することができ、ユーザーの共通の話題を抽出することができ、ある時点からある時点ま

*ができる。

【0068】図11は、本発明の第3の実施の形態を示している。本実施の形態では、第1の実施の形態で生成された文書情報記憶部16に格納された知識データを用い、2つ以上の対象区間同士を比較することにより、ユーザーの文書情報(ユーザーの興味)の推移を表わす機能を付加している。係る推移を求めるのが、ユーザーの興味推移算出部19であり、求めた結果を出力部18に出力する。

【0069】本実施の形態の意図を理解しやすくするために、具体例をあげつつ説明する。まず、ユーザーの先月と今月の興味を表わすキーワードが、以下のようになっているものとする。

【0070】

【表2】

で、現れなかった目新しいキーワードを抽出し、それをもとに、目新しい文書を抽出することができる等のユーザーの推移を表わすことができる。

【0074】上記した各実施の形態は、いずれも装置に組み込んだ状態を前提として説明したが、上記した各処理をコンピュータに実行させるためのプログラムとし、係るプログラムは所定の記録媒体に記録して提供するようにしてもよい。一例を示すと、図3、図4、図7、図11に示すフローチャートの各処理ステップを実行するためのプログラムであったり、ユーザーの興味推移算出を行うプログラムであったりする。そして、上記した記録媒体に記録するプログラムは、各フローチャートをすべて実行するためのプログラムではなく、例えば図3のフローチャートを実行するためのプログラムのみというように、一部でもよい。

【0075】図13、図14に示すように、記録媒体としては、フロッピーディスク(FD)30やCD-ROM31などがあり、係る記録媒体30、31に格納されたプログラムは、FDドライブ32やCD-ROMドライブ33を介してコンピュータ34に接続された(内蔵された)HDユニット35にインストールされ、これにより、コンピュータ34は、上記した各実施の形態で説明した装置を構成することになる。具体的には、各実施の形態における記憶部13、16、24などは、HDユニット35により実現され、各抽出部12、15、22や評価部17、23、知識更新部25、ユーザーの興味推移算出部19は、例えば使用の都度HDユニット35から内部メモリ36に転送され、そこにおいて各種処理を高速に行うことができる。また、文書情報の入力部1

19

1, 21は、キーボード37やモデム38等により実現され、時間キー入力部14はキーボード37等により実現される。さらに、出力部18, 26は、図示の例ではCRTディスプレイ（表示装置）39に対応するが、プリンタその他各種の出力装置を用いることができる。

【0076】

【発明の効果】以上のように、本発明に係る文書処理方法及び装置並びに記録媒体では、従来の文書特徴量に加え、時間的要素を加味することにより、ユーザーの興味対象等が変化してもそれに追従して知識データを変更し、精度のよい判断を行うことができ、また、ユーザーの興味対象やその逆が時間とともにどのように推移していったかを検出することができる。

【図面の簡単な説明】

【図1】本発明の第1の実施の形態の概念を説明する図である。

【図2】本発明の第1の実施の形態を示すブロック図である。

【図3】第1の実施の形態の装置の動作の一部を説明するフローチャートである。

【図4】第1の実施の形態の装置の動作の一部を説明するフローチャートである。

【図5】文書情報とキーワード集合の記憶部13のデータ構造と登録されたデータの一例を示す図である。

【図6】対象区間内の文書情報記憶部16のデータ構造と登録されたデータの一例を示す図である。

【図7】第1の実施の形態の装置の動作の一部を説明するフローチャートである。

【図8】本発明の第1の実施の形態を示すブロック図である。

【図9】文書情報記憶部24のデータ構造と登録された

20

データの一例を示す図で、(A)が更新前で(B)が更新後のデータをそれぞれ示す図である。

【図10】第2の実施の形態の装置の動作を説明するフローチャートである。

【図11】本発明の第3の実施の形態を示すブロック図である。

【図12】第3の実施の形態におけるユーザーの興味推移算出部の機能を説明する図である。

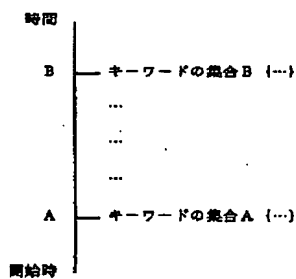
【図13】本発明に係る記録媒体を実施するシステム構成を示す図である。

【図14】本発明に係る記録媒体を実施するシステム構成を示す図である。

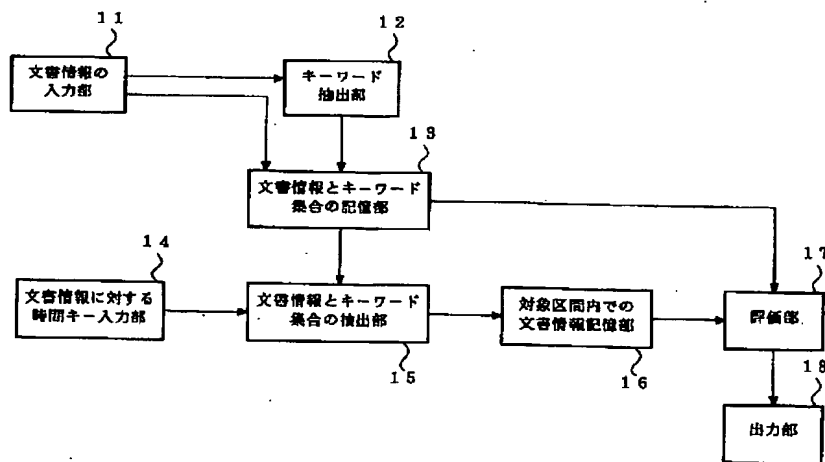
【符号の説明】

- 11 文書情報の入力部
- 12 キーワード抽出部
- 13 文書情報とキーワード集合の記憶部
- 14 文書情報に対する時間キー入力部
- 15 文書情報とキーワード集合の抽出部
- 16 対象区間内の文書情報記憶部
- 17 評価部
- 18 出力部
- 19 ユーザーの興味推移算出部
- 21 文書情報の入力部
- 22 キーワード抽出部
- 23 評価部
- 24 文書情報記憶部
- 25 知識更新部
- 26 出力部
- 30 FD（記録媒体）
- 31 CD-ROM（記録媒体）

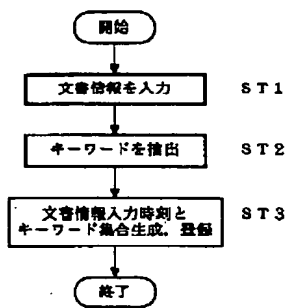
【図1】



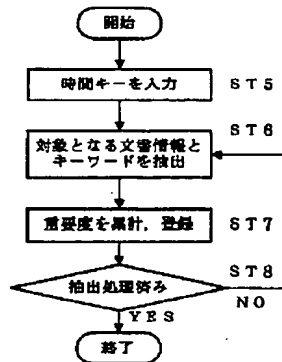
【図2】



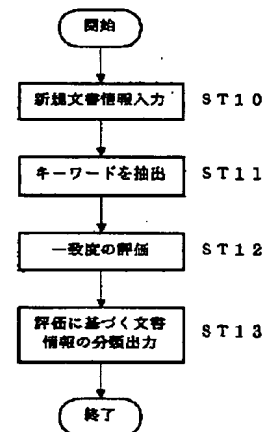
【図3】



【図4】



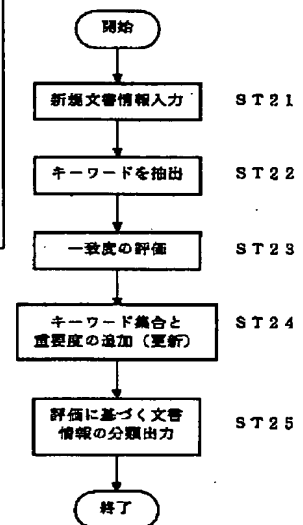
【図7】



【図5】

文書名	時間	キーワード (重要度)
A1	1996/05/01 12:50	0157(0.9), 柳市(0.8), 食中毒(0.7), 給食(0.3), ...
...
D1	1997/05/12 01:10	ペルー(0.8), 人質(0.5), フジモリ(0.6), 青木(0.2), ...
...
D13	1997/05/30 00:00	ペルー(0.6), 青木(0.7), 橋本(0.8), 毎田(0.4), ...
...

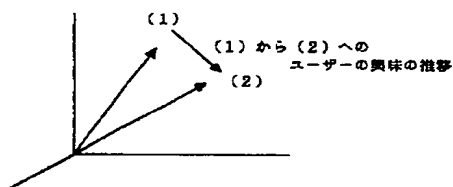
【図10】



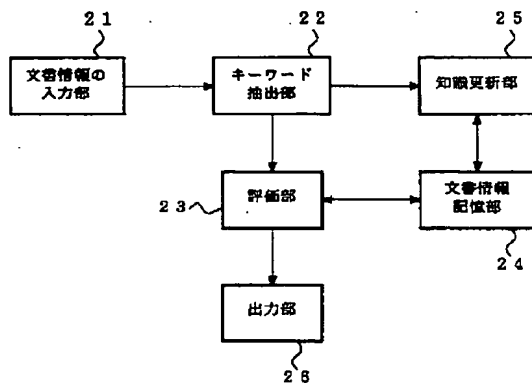
【図6】

時間	キーワード (重要度)
1997/05/12 01:10~ 1997/05/30 00:00	ペルー(0.7), 人質(0.6), フジモリ(0.3), 青木(0.1), ...

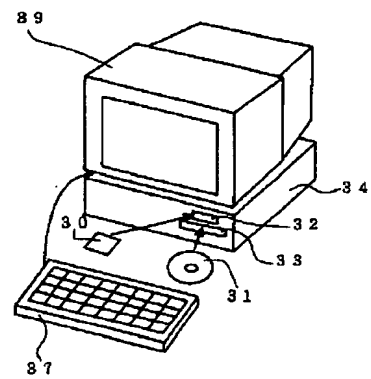
【図12】



【図8】



【図13】

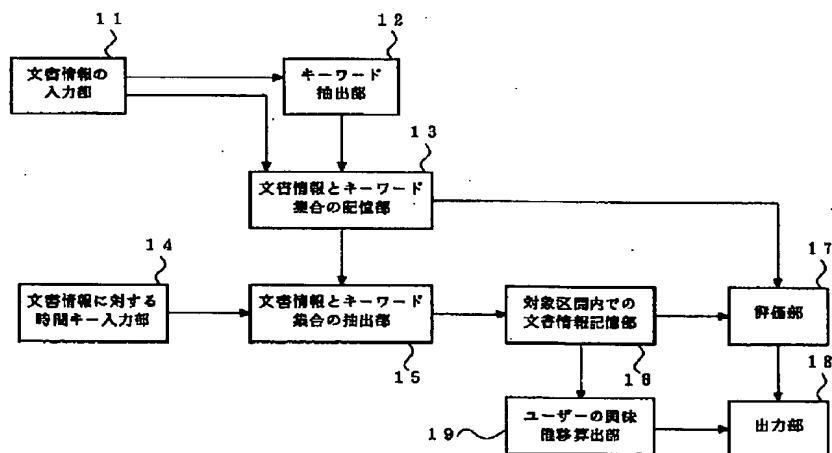


【図9】

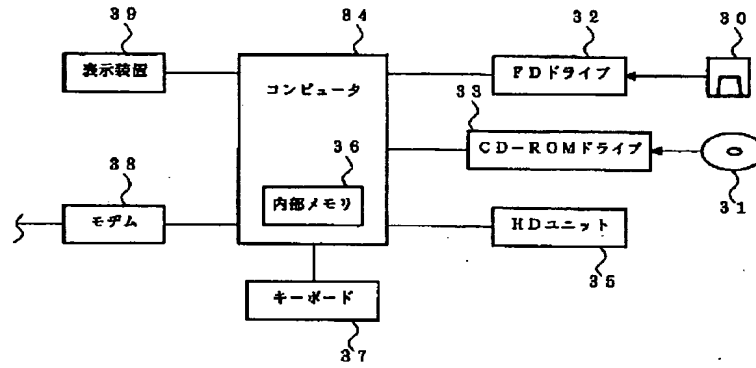
	時 間	キーワード(重要度)	...
(A)	1997/05/12 01:10	ペルー(0.8), 人質(0.5), フジモリ(0.6), 青木(0.2),

	時 間	キーワード(重要度)	...
(B)	1997/05/12 01:10	ペルー(0.6), 人質(0.2), フジモリ(0.3), 青木(0.5), 橋本(0.4), 帰国(0.2),

【図11】



【図14】



フロントページの続き

(72)発明者 呉 亜棟

京都府京都市右京区花園土堂町10番地 オ
ムロン株式会社内